Self-simplification in Darwin's Systems

Alexander N. Gorban

University of Leicester, LE1 7RH Leicester, UK, ag153@le.ac.uk

Abstract. We prove that a non-linear kinetic system with *conservation of supports* for distributions has generically limit distributions with final support only. The conservation of support has a biological interpretation: *inheritance*. We call systems with inheritance "Darwin's systems". Such systems are apparent in many areas of biology, physics (the theory of parametric wave interaction), chemistry and economics. The finite dimension of limit distributions demonstrates effects of *natural selection*. Estimations of the asymptotic dimension are presented. After some initial time, solution of a kinetic equation with conservation of support becomes a finite set of narrow peaks that become increasingly narrow over time and move increasingly slowly. It is possible that these peaks do not tend to fixed positions, and the path covered tends to infinity as $t \to \infty$. The *drift equations* for peak motion are obtained. They describe the asymptotic layer near the *omega*-limit distributions with finite support .

1 Introduction: Unusual Conservation Law

How can we prove that all the attractors of an infinite-dimensional system belong to a finite-dimensional manifold? How can we estimate the dimensions of attractor? There exist two methods to obtain such estimations.

First, if we find that *k-dimensional volumes are contracted* due to dynamics, then (after some additional technical steps) we can claim that the *Hausdorff dimension* of the maximal attractor is less than k. This idea is in the essence of *inertial manifold* theory [11]. The standard way to prove this k-dimensional volumes contraction is to check that the symmetrized Jakobian operator has discrete spectrum and the sum of any k its eigenvalues has negative real part (under some additional conditions of uniform boundness of solutions).

Second, if we find a representation of our system as a nonlinear kinetic system with *conservation of supports* of distributions, then (again, after some additional technical steps) we can state that the asymptotics is finite-dimensional: the distribution evolves in a sum of several narrow peaks of density. This conservation of support has a *quasi-biological interpretation, inheritance* (if a gene was not presented initially in an isolated population without mutations, then it cannot appear at later time). The finite-dimensional asymptotic demonstrates effects of *"natural" selection*. It is very natural to call them *Darwin's systems.*

In the 1970s to the 1980s, theoretical work developed another "common" field simultaneously applicable to physics, biology and mathematics.

For physics it is (so far) part of the theory of a special kind of approximation, demonstrating, in particular, interesting mechanisms of discreteness in the course of the evolution of distributions with initially smooth densities. However, what for physics is merely a convenient approximation is a fundamental law in biology: inheritance. The consequences of inheritance (collected in the selection theory [13,27,41,48,49,3,17,28,21,25,47]) give one of the most important tools for biological reasoning.

Consider a community of animals. Let it be biologically isolated. Mutations can be neglected in the first approximation. In this case, new genes do not emerge. Support of the distribution of genes does not increase.

An example from physics is also possible and leads to a very seminal approach to nonlinear wave theory. Let waves with wave vectors k be excited in some system. Denote K a set of wave vectors k of excited waves. Let the wave interaction does not lead to the generation of waves with new $k \notin K$. Such an approximation is applicable to a variety of situations, and has been described in detail for wave turbulence in [53,54].

What is common in these examples is the evolution of a distribution with a support that does not increase over time.

What does not increase must, as a rule, decrease, if the decrease is not prohibited. This naive thesis can be converted into rigorous theorems for the case under consideration [21]. It is proved that the support decreases in the limit $t \to \infty$ if it was sufficiently large initially. (At finite times the distribution supports are conserved and decrease only in the limit $t \to \infty$.) Conservation of the support usually results in the following effect: dynamics of an initially infinite-dimensional system at $t \to \infty$ can be described by finite-dimensional systems and distribution degenrates into a sum of finite number of narrow peaks, the peaks' width tends to zero and the ω-limit distributions are finite sums of δ-functions.

This is description of the final, ω-limit distribution. More precisely, it is given by the selection theorem, and the dimension of the limiting systems can be evaluated by the properties of the reproduction coefficient functions.

In this paper we are focused on another problem: how a Darwin system approaches this finite-dimensional asymptotic? The first naive expectation is as follows: dynamics of infinite system tends to a finite-dimensional dynamics, which is predefined by the support of initial distribution and the adaptation landscape. This is, on some sense, another example of finite-dimensional inertial manifold.

Such asymptotic behavior of Darwin systems is possible, but surprisingly it was demonstrated [46,20,21] that even in rather simple examples Darwin's systems do not obligatory tend to one finite-dimensional asymptotic dynamical system, but can wander near infinitely many such systems. This wandering becomes slower in time, the density peaks become narrower, but the way of the wandering may tend to infinity because the velocity does not tends to zero in logarithmic time, $\log t$. We call this wandering the "drift effect". Below we describe the asymptotic behavior of systems with inheritance approaching their ω-limit distributions.

The simplest model for "reproduction + small mutations" where mutations are presented as diffusion is also studied. The limit of zero mutations is singular, and in the systems with small mutations the zero limit of the drift velocity of their drift at $t \to \infty$ substitutes by a small finite one in the presence of drift effect. Moreover, there exists a *scale invariance*, and dynamics for large t does not depend on nonzero mutation intensity, if the last is sufficiently small: to change this intensity, we need just to rescale time.

The structure of the paper is as follows. In Sec. 2 Darwin's systems are formally described and selection theorems are presented. In Sec. 2.2 the optimality principles for supports of ω-limit distributions are developed. These principles have a "weak" form; the set of possible supports is estimated from above and it is not obvious that this estimation is effective. A theorem of selection efficiency is presented in Sec. 2.5. The sense of this theorem is as follows: for almost every system the support of all ω-limit distributions is small (in an appropriate strong sense). Its geometrical interpretation suggested by M. Gromov is explained in Sec. 2.6.

Minimax estimations of the number of points in the support of ω-limit distributions are given in Section 2.3. The idea is to study systems under a ε-small perturbation, to estimate the maximal number of points for each realization of the perturbed system, and then to estimate the minimum of these maxima among various realizations. These minimax estimates can be constructive and do not use integration of the system. The set of reproduction coefficients $\{k(\mu) \,|\, \mu \in M\}$ is compact in $C(X)$. Therefore, this set can be approximated by a finite–dimensional linear space L_ε with any given accuracy ε.

The number of coexisting inherited units ("quasi-species") is estimated from above as $\dim L_\varepsilon$. This estimate is true both for stationary and nonstationary coexistence. In its general form this estimate was proved in 1980 [20,21], but the reasoning of this type has a long history. Perhaps, G. Gause [19] was the first to suggest the direct connection between the number of species and the number of resources. One can call this number "dimension of the environment." He proposed the famous concurrent exclusion principle.

More details about early history of the concurrent exclusion principle are presented in the review paper of G. Hardin [29].

MacArthur and Levins [39] suggested that the number of coexisting species is limited by the number of ecological resources. Later [40], they studied the continuous resource distribution (niche space) where the number of species is limited by the fact that the niches must not overlap too much. In 1999, G. Meszena and J.A.J. Metz [43] developed further the idea of environmental feedback dimensionality (perhaps, independently of [20,21]).

In 2006 the idea of robustness in concurrent exclusion was approached again, as a "unified theory" of "competitive exclusion and limiting similarity" [42]. All these achievements are related to estimation of dimension of the set

$\{k(\mu) \mid \mu \in M\}$ or of some its subsets. This dimension plays the role of "robust dimension of population regulation".

Drift equations without mutations are derived in Sec. 3.1 and asymptotic equations For Darwin's systems with small mutations with scaling invariance theorem are presented in Sec. 3.2.

The special structure of Darwin's systems requires to introduce three types of stability of the limit behavior (Sec. 4): internal stability (stability in the limiting finite-dimensional systems), external stability or uninvadability (stability with respect to strongly small perturbations that extend the support), and stable realizability (stability with respect to small shifts and extensions of the density peaks). Internal stability (Sec. 4.1) is just a particular case of Lyapunov stability applied to finite-dimensional asymptotic. External stability (uninvadability, Sec. 4.2) was first introduced and studied by J.B.S. Haldane [27], then, after papers [48,49] it was intensively used in biology [50,31,**?**,51,12,9,4,5,8,6,7], and general evolutionary games theory [15,44,45,32]. In physics, this notion was introduced in an entirely independent series of works on the S-approximation in the spin wave theory and on wave turbulence [53,54,38], which studied wave configurations in the approximation of an "inherited" wave vector.

The stable realizability (stability with respect to small shifts and extensions of the density peaks, Sec. 4.3) was first introduced and studied in full generality in early 1980th [21]. Later, the important particular cases were independently introduced and studied in series of papers [44,45,15]. In these papers the idea of drift equations appeared for the gaussian peaks in the dynamics of continuous symmetric evolutionary games. The authors [44,45] called this property of "stable realizability" by "evolutionary robustness" and claimed the necessity of this additional type of stability very energetically: "Furthermore, we provide new conditions for the stability of rest points and show that even strict equilibria may be unstable".

In Sec. 5 a rich family of examples is described. Those examples are the generalized Lottka–Volterra–Gause infinite dimensional systems with distributed coefficients. The main benefit from this special structure is the generalized Volterra averaging principle [52]. This principle allows to substitute the time averages of linear functionals by their values at steady states. For the distributed Lottka–Volterra–Gause systems the drift equations are written explicitly.

In Section 7 a brief description of the main results is presented.

2 Inheritance and Selection Theorems

2.1 Asymptotically Stable States

The simplest and most common class of equations in applications for which the distribution support does not grow over time is constructed as follows.

To each distribution μ is assigned a function k_μ by which distributions can be multiplied. Let us write down the equation:

$$\frac{\mathrm{d}\mu}{\mathrm{d}t} = k_\mu \times \mu\,. \tag{1}$$

The multiplier k_μ is called a *reproduction coefficient*. It depends on μ, and this dependence can be rather general and non-linear.

Two remarks can be important:

1. The apparently simple form of (1) does not mean that this system is linear or even close to linear. The operator $\mu \mapsto k_\mu$ is a general non-linear operator, and the only restriction is its continuity in an appropriate sense (see below).
2. On a finite set $X = \{x_1, \dots, x_n\}$, non-negative measures μ are simply non-negative vectors $\mu_i \geq 0$ (i=1,..., n), and (1) appears to be a system of equations of the following type:

$$\frac{\mathrm{d}\mu_i}{\mathrm{d}t} = k_i(\mu_1, \dots, \mu_n) \times \mu_i\ , \tag{2}$$

and the only difference from a general dynamic system is the special behavior of the right-hand side of (2) near zero values of μ_i.

The right-hand side of (1) is the product of the function k_μ and the distribution μ, and hence $\mathrm{d}\mu/\mathrm{d}t$ should be zero when μ is equal to zero; therefore the support of μ is conserved in time (over finite times).

X the *space of inherited units*. Most of the selection theorems are proved for compact metric space X with a metric $\rho(x, y)$. Further, for the drift and mutations equations we assume that X is a domain in finite-dimensional real space R^n (closed and bounded). As a particular case of compact space, a finite set X can be discussed.

μ is distribution on X. Each distribution on a compact space X is a continuous linear functional on the space of continuous real functions $C(X)$. We follow the Bourbaki approach [10]: a measure is a continuous functional, an integral. Book [10] contains all the necessary notions and theorems. Space $C(X)$ is a Banach space endowed with the maximum norm

$$\|f\| = \max_{x \in X} |f(x)|\ . \tag{3}$$

If $\mu \in C^*(X)$ and $f \in C(X)$, then $[\mu, f]$ is the value of μ at a function f. If X is a bounded closed subset of a finite-dimensional space R^n, then we represent this functional as the integral

$$[\mu, f] = \int \mu(x) f(x)\,\mathrm{d}x\ , \tag{4}$$

which is the standard notation for distribution (or generalized function) theory. The "density" $\mu(x)$ is not assumed to be an absolute continuous function with respect to the Lebesgue measure dx (or even a function), and the notation in Eq. (4) has the same sense as $[\mu, f]$. If the measure is defined as a function on a σ-algebra of sets, then the following notation is used:

$$[\mu, f] = \int f(x)\, \mu(\mathrm{d}x) \ .$$

We use the notation $[\mu, f]$ for general spaces X and the representation (4) on domains in R^n. The product $k \times \mu$ is defined for any $k \in C(X)$, $\mu \in C^*(X)$ by the equality: $[k\mu, f] = [\mu, kf]$.

The support of μ, suppμ, is the smallest closed subset of X with the following property: if $f(x) = 0$ on suppμ, then $[\mu, f] = 0$, i.e. $\mu(x) = 0$ outside suppμ.

In the space of measures we use *weak* convergence*, i.e. the convergence of averages:

$$\mu_i \to \mu^* \text{ if and only if } [\mu_i, \varphi] \to [\mu^*, \varphi] \tag{5}$$

for all continuous functions $\varphi \in C(X)$. This weak* convergence of measures generates *weak* topology* on the space of measures (sometimes called weak topology of conjugated space, or wide topology).

Strong topology on the space of measures $C^*(X)$ is defined by the norm $\|\mu\| = sup_{\|f\|=1}[\mu, f]$.

The properties of the mapping $\mu \mapsto k_\mu$ should be specified, and the existence and uniqueness of solutions of (1) under given initial conditions should be identified. In specific situations the answers to these questions are not difficult.

The sequence of continuous functions $k_i(x)$ is considered to be convergent if it converges uniformly. The sequence of measures μ_i is called convergent if for any continuous function φ the integrals $[\mu_i, \varphi]$ converge [weak* convergence (5)]. The mapping $\mu \mapsto k_\mu$ assigning the reproduction coefficient k_μ to the measure μ is assumed to be continuous with respect to these convergencies.

The space of measures is assumed to have a bounded set M that is positively invariant relative to system (1): if $\mu(0) \in M$, then $\mu(t) \in M$ (we also assume that M is non-trivial, i.e. it is neither empty nor a one-point set i.e. it is neither empty nor a one-point set but includes at least one point with its vicinity). This M serves as the phase space of system (1). (Let us remind that the set of measures M is bounded if the set of integrals $\{[\mu, f] \,|\, \mu \in M, \|f\| \leq 1\}$ is bounded, where $\|f\|$ is the norm (3).) We study dynamic of system (1) in bounded positively invariant set M.

Most of the results for systems with inheritance use a **theorem on weak* compactness**: *The bounded set of measures is precompact with respect to weak* convergence (i.e. its closure is compact).* Therefore the set of corresponding reproduction coefficients $k_M = \{k_\mu \,|\, \mu \in M\}$ is precompact.

The simplest example of an emerging discrete distribution from a continuous initial distribution gives us the following equation:

$$\frac{\partial\mu(x,t)}{\partial t} = \left[f_0(x) - \int_a^b f_1(x)\mu(x,t)\,\mathrm{d}x \right] \mu(x,t) \ , \tag{6}$$

where the functions $f_0(x)$ and $f_1(x)$ are positive and continuous on the closed segment $[a,b]$. Let the function $f_0(x)$ reach the global maximum on the segment $[a,b]$ at a single point x_0. If $x_0 \in \mathrm{supp}\mu(x,0)$, then:

$$\mu(x,t) \to \frac{f_0(x_0)}{f_1(x_0)}\delta(x-x_0), \qquad \text{when } t \to \infty \ , \tag{7}$$

where $\delta(x-x_0)$ is the δ-function.

If $f_0(x)$ has several global maxima, then the right-hand side of (7) can be the sum of a finite number of δ-functions. Here a natural question arises: is it worth considering such a possibility? Indeed, such a case seems to be very unlikely to occur. More details on this are given below.

The limit behavior of a typical system with inheritance (1) can be much more complicated than (7). Here we can mention that any finite-dimensional system with a compact phase space can be embedded in a system with inheritance (2). An additional possibility for the limit behavior is, for example, the drift effect (Section 3.1).

The first step in the routine investigation of a dynamical system is a question about fixed points and their stability. The first observation concerning the system (1) is that it can only be asymptotically stable for steady-state distributions, the support of which is discrete (i.e. the sums of δ-functions). This can be proved for all consistent formalizations.Thus, we have the first theorem.

Theorem 1. *The support of asymptotically stable distributions for the system (1) is always discrete.* □

For the proof of this theorem and other selection theorems we refer to [20,21,23,24]

The perturbation discussed is small not only in the weak* topology, but also in the strong sense, and thus it is sufficient to consider strongly small perturbations to prove that the asymptotically stable distribution should be discrete. Hence, this statement is true if the operator $\mu \mapsto k_\mu$ is continuous for strong topology on the space of measures. This is a significantly weaker requirement than being continuous in weak* topology.

This simple observation has many strong generalizations to general ω-limit points, to equations for vector measures, etc.

2.2 Optimality Principle for Limit Diversity

Description of the limit behavior of a dynamical system can be much more complicated than enumerating stable fixed points and limit cycles. The leading rival to adequately formalize the limit behavior is the concept of the "ω-limit set". It was discussed in detail in the classical monograph [1]. The fundamental textbook on dynamical systems [30] and the introductory review [34] are also available.

Let $f(t)$ be the dependence of the position of point in the phase space on time t (i.e. the *motion* of the dynamical system). A point y is a ω-limit point of the motion $f(t)$, if there exists such a sequence of times $t_i \to \infty$, that $f(t_i) \to y$.

The set of all ω-limit points for the given motion $f(t)$ is called the ω-limit set. If, for example, $f(t)$ tends to the equilibrium point y^* then the corresponding ω-limit set consists of this equilibrium point. If $f(t)$ is winding onto a closed trajectory (the limit cycle), then the corresponding ω-limit set consists of the points of the cycle and so on.

General ω-limit sets are not encountered oft in specific situations. This is because of the lack of efficient methods to find them in a general situation. Systems with inheritance is a case, where there are efficient methods to estimate the limit sets from above. This is done by the optimality principle.

Let $\mu(t)$ be a solution of (1). Note that

$$\mu(t) = \mu(0) \exp \int_0^t k_{\mu(\tau)} \, \mathrm{d}\tau \ . \tag{8}$$

Here and below we do not display the dependence of distributions μ and of the reproduction coefficients k on x when it is not necessary. Fix the notation for the average value of $k_{\mu(\tau)}$ on the segment $[0, t]$

$$\langle k_{\mu(t)} \rangle_t = \frac{1}{t} \int_0^t k_{\mu(\tau)} \, \mathrm{d}\tau \ . \tag{9}$$

Then the expression (8) can be rewritten as

$$\mu(t) = \mu(0) \exp(t \langle k_{\mu(t)} \rangle_t) \ .$$

If μ^* is the ω-limit point of the solution $\mu(t)$, then there exists such a sequence of times $t_i \to \infty$, that $\mu(t_i) \to \mu^*$. Let it be possible to chose a convergent subsequence of the sequence of the average reproduction coefficients $\langle k_{\mu(t)} \rangle_t$, which corresponds to times t_i. We denote as k^* the limit of this subsequence. Then, the following statement is valid: on the support of μ^* the function k^* vanishes and on the support of $\mu(0)$ it is non-positive:

$$\begin{aligned} &k^*(x) = 0 \text{ if } x \in \mathrm{supp}\mu^*, \\ &k^*(x) \leq 0 \text{ if } x \in \mathrm{supp}\mu(0) \ . \end{aligned} \tag{10}$$

Taking into account the fact that supp$\mu^* \subseteq$ supp$\mu(0)$, we come to the formulation of **the optimality principle** (10): *The support of limit distribution consists of points of the global maximum of the average reproduction coefficient on the initial distribution support. The corresponding maximum value is zero.*

We should also note that not necessarily all points of maximum of k^* on supp$\mu(0)$ belong to suppμ^*, but all points of suppμ^* are the points of maximum of k^* on supp$\mu(0)$.

If $\mu(t)$ tends to the fixed point μ^*, then $\langle k_{\mu(t)} \rangle_t \to k_{\mu^*}$ as $t \to \infty$, and suppμ^* consists of the points of the global maximum of the corresponding reproduction coefficient k_{μ^*} on the support of μ^*. The corresponding maximum value is zero.

If $\mu(t)$ tends to the limit cycle $\mu^*(t)$ $(\mu^*(t+T) = \mu^*(t))$, then all the distributions $\mu^*(t)$ have the same support. The points of this support are the points of maximum (global, zero) of the averaged over the cycle reproduction coefficient

$$k^* = \langle k_{\mu^*(t)} \rangle_T = \frac{1}{T} \int_0^T k_{\mu^*(\tau)} \, \mathrm{d}\tau \ ,$$

on the support of $\mu(0)$.

The supports of the ω-limit distributions are specified by the functions k^*. It is obvious where to get these functions from for the cases of fixed points and limit cycles. There are at least two questions: what ensures the existence of average reproduction coefficients at $t \to \infty$, and how to use the described extremal principle (and how efficient is it). The latter question is the subject to be considered in the following sections. In the situation to follow the answers to these questions have the validity of theorems.

Due to the theorem about weak* compactness, the set of reproduction coefficients $k_M = \{k_\mu \,|\, \mu \in M\}$ is precompact, hence, the set of averages (9) is precompact too, because it is the subset of the closed convex hull $\overline{\mathrm{conv}}(k_M)$ of the compact set. This compactness allows us to claim the existence of the *average reproduction coefficient* k^* for the description of the ω-limit distribution μ^* with the optimality principle (10).

2.3 How Many Points Does the Limit Distribution Support Hold?

The limit distribution is concentrated in the points of (zero) global maximum of the average reproduction coefficient. The average is taken along the solution, but the solution is not known beforehand. With the convergence towards a fixed point or to a limit cycle this difficulty can be circumvented. In the general case the extremal principle can be used without knowing the solution, in the following way [21]. Considered is a set of all dependencies $\mu(t)$ where μ belongs to the phase space, the bounded set M. The set of all averages over t is $\{\langle k_{\mu(t)} \rangle_t\}$. Further, taken are all limits of sequences formed by these averages – the set of averages is closed. The result is the closed convex hull $\overline{\text{conv}}(k_M)$ of the compact set k_M. This set involves all possible averages (9) and all their limits. In order to construct it, the true solution $\mu(t)$ is not needed.

The weak optimality principle is expressed as follows. Let $\mu(t)$ be a solution of (1) in M, μ^* is any of its ω-limit distributions. Then in the set $\overline{\text{conv}}(k_M)$ there is such a function k^* that its maximum value on the support $\text{supp}\mu_0$ of the initial distribution μ_0 equals to zero, and $\text{supp}\mu^*$ consists of the points of the global maximum of k^* on $\text{supp}\mu_0$ only (10).

Of course, in the set $\overline{\text{conv}}(k_M)$ there are usually many functions that are irrelevant to the time average reproduction coefficients for the given motion $\mu(t)$. Therefore, the weak extremal principle is really weak – it gives too many possible supports of μ^*. However, even such a principle can help to obtain useful estimates of the number of points in the supports of ω-limit distributions.

It is not difficult to suggest systems of the form (1), in which any set can be the limit distribution support. The simplest example: $k_\mu \equiv 0$. Here ω-limit (fixed) is any distribution. However, almost any arbitrary small perturbation of the system destroys this pathological property.

In the realistic systems, especially in biology, the coefficients fluctuate and are never known exactly. Moreover, the models are in advance known to have a finite error which cannot be exterminated by the choice of the parameters values. This gives rise to an idea to consider not individual systems (1), but ensembles of similar systems [21].

Let us estimate the maximum for each individual system from the ensemble (in its ω-limit distributions), and then, estimate the minimum of these maxima over the whole ensemble – (*the minimax estimation*). The latter is motivated by the fact, that if the inherited unit has gone extinct under some conditions, it will not appear even under the change of conditions.

Let us consider an ensemble that is simply the ε-neighborhood of the given system (1). The minimax estimates of the number of points in the support of ω-limit distribution are constructed by approximating the dependencies k_μ by finite sums

$$k_\mu = \varphi_0(x) + \sum_{i=1}^{n} \varphi_i(x)\psi_i(\mu) \ . \tag{11}$$

Here φ_i depend on x only, and ψ_i depend on μ only. Let $\varepsilon_n > 0$ be the distance from k_μ to the nearest sum (11) (the "distance" is understood in the suitable rigorous sense, which depends on the specific problem). So, we reduced the problem to the estimation of the diameters $\varepsilon_n > 0$ of the set $\overline{\text{conv}}(k_M)$.

The minimax estimation of the number of points in the limit distribution support gives the answer to the question, "How many points does the limit distribution support hold": *If $\varepsilon > \varepsilon_n$ then, in the ε-vicinity of k_μ, the minimum of the maxima of the number of points in the ω-limit distribution support does not exceed n.*

In order to understand this estimate it is sufficient to consider system (1) with k_μ of the form (11). In this case for any dependence $\mu(t)$ the averages (9) have the form

$$\langle k_{\mu(t)}\rangle_t = \frac{1}{t}\int_0^t k_{\mu(\tau)}\,\mathrm{d}\tau = \varphi_0(x) + \sum_{i=1}^{n}\varphi_i(x)a_i \ . \qquad (12)$$

where a_i are some numbers. The ensemble of the functions (12) for various a_i forms a n-dimensional linear manifold. How many points of the global maximum (equal to zero) could a function of this family have?

Generally speaking, it can have any number of maxima. However, it seems obvious, that "usually" one function has only one point of global maximum, while it is "improbable" that the maximum value is zero. At least, with an arbitrary small perturbation of the given function, we can achieve for the point of the global maximum to be unique and the maximum value be non-zero.

In a one-parametric family of functions there may occur zero value of the global maximum, which cannot be eliminated by a small perturbation, and individual functions of the family may have two global maxima.

In the general case we can state, that "usually" each function of the n-parametric family (12) can have not more than n points of the zero global maximum (of course, there may be less, and the global maximum is, as a rule, not equal to zero at all for the majority of functions of the family). What "usually" means here requires a special explanation given in the next section.

In application k_μ is often represented by an integral operator, linear or nonlinear. In this case the form (11) corresponds to the kernels of integral operators, represented in a form of the sums of functions' products. For example, the reproduction coefficient of the following form

$$k_\mu = \varphi_0(x) + \int K(x,y)\mu(y)\,\mathrm{d}y \ ,$$
$$\text{where } K(x,y) = \sum_{i=1}^{n}\varphi_i(x)g_i(y) \ , \qquad (13)$$

has also the form (11) with $\psi_i(\mu) = \int g_i(y)\mu(y)\,\mathrm{d}y$.

The linear reproduction coefficients occur in applications rather frequently. For them the problem of the minimax estimation of the number of points in the ω-limit distribution support is reduced to the question of the accuracy of approximation of the linear integral operator by the sums of kernels-products (13).

2.4 Almost Finite Sets and "Almost Always"

The supports of the ω-limit distributions for the systems with inheritance were characterized by the optimality principle. These supports consist of points of global maximum of the average reproduction coefficient. We can a priori (without studying the solutions in details) characterize the compact set that includes all possible average reproduction coefficients. Hence, we get a problem: how to describe the set of global maximum for all functions from generic compact set of functions. First of all, any closed subset $M \subset X$ is a set of global maximum of a continuous function, for example, of the function $f(x) = -\rho(x, M)$, where $\rho(x, M)$ is the distance between a set and a point: $\rho(x, M) = \inf_{y \in M} \rho(x, y)$, and $\rho(x, y)$ is the distance between points. Nevertheless, we can expect that one generic function has one point of global maximum, in a generic one-parametric family might exist functions with two points of global maximum, etc. How these expectations meet the exact results? What does the notion "generic" mean? What can we say about sets of global maximum of functions from a generic compact family? In this section we answer these questions.

Here are some examples of correct but useless statements about "generic" properties of function: Almost every continuous function is not differentiable; Almost every C^1-function is not convex. Their meaning for applications is most probably this: the genericity used above for continuous functions or for C^1-function is irrelevant to the subject.

Most frequently the motivation for definitions of genericity is found in such a situation: given n equations with m unknowns, what can we say about the solutions? The answer is: in a typical situation, if there are more equations, than the unknowns ($n > m$), there are no solutions at all, but if $n \leq m$ (n is less or equal to m), then, either there is a $(m - n)$-parametric family of solutions, or there are no solutions.

The best known example of using this reasoning is the *Gibbs phase rule* in classical chemical thermodynamics. It limits the number of co-existing phases. There exists a well-known example of such reasoning in mathematical biophysics too. Let us consider a medium where n species coexist. The medium is assumed to be described by m parameters s_j. Dynamics of these parameters depends on the organisms. In the simplest case, the medium is a well-mixed solution of m substances. Let the organisms interact through the medium, changing its parameters – concentrations of m substances. It can be formalized by a system of equation:

$$\frac{\mathrm{d}\mu_i}{\mathrm{d}t} = k_i(s_1, \ldots, s_m) \times \mu_i \, (i = 1, \ldots n) \; ;$$
$$\frac{\mathrm{d}s_j}{\mathrm{d}t} = q_j(s_1, \ldots, s_m, \mu_1, \ldots, \mu_n) \, (j = 1, \ldots m) \; , \tag{14}$$

In a steady state, for each of the coexisting species we have an equation with respect to the state of the medium: the corresponding reproduction coefficient k_i is zero. So, the number of such species cannot exceed the number of parameters of the medium. In a typical situation, in the m-parametric medium in a steady state there can exist not more than m species. This is the concurrent exclusion principle in the G. Gause form [19]. Here, the main hypothesis about interaction of organisms with the media is that the number of essential components of the media is bounded from above by m and increase of the number of species does not extend the list of components further. Dynamics of parameters depends on the organisms, but their nomenclature is fixed.

This concurrent exclusion principle allows numerous generalizations [39,40,37,14,42]. Theorem of the natural selection efficiency may be also considered as its generalization.

Analogous assertion for a non-steady state coexistence of species in the case of equations (14) is not true. It is not difficult to give an example of stable coexistence under oscillating conditions of n species in the m-parametric medium at $n > m$.

But, if $k_i(s_1, \ldots, s_m)$ are linear functions of $s_1, \ldots, s_m$, then for non-stable conditions we have the concurrent exclusion principle, too. In that case, the average in time of the reproduction coefficient is the reproduction coefficient for the average state of the medium:

$$\langle k_i(s_1(t), \ldots, s_m(t))\rangle = k_i(\langle s_1\rangle, \ldots, \langle s_m\rangle)$$

because of linearity. If $\langle x_i\rangle \neq 0$ then $k_i(\langle s_1\rangle, \ldots, \langle s_m\rangle) = 0$, and we obtain the non-stationary concurrent exclusion principle "in average". And again, it is valid "almost always".

The non-stationary concurrent exclusion principle "in average" is valid for linear reproduction coefficients. This is a combination of the Volterra [52] averaging principle and the Gause principle,

It is worth to mention that, for our basic system (1), if k_μ are linear functions of μ, then the average in time of the reproduction coefficient $k_{\mu(t)}$ is the reproduction coefficient for the average $\mu(t)$ because of linearity. Therefore, the optimality principle (10) for the average reproduction coefficient k^*, transforms into the following optimality principle for the reproduction coefficient $k_{\langle\mu\rangle}$ of the average distribution $\langle\mu\rangle$

$$\begin{aligned} &k_{\langle\mu\rangle}(x) = 0 \text{ if } x \in \mathrm{supp}\mu^* \ , \\ &k_{\langle\mu\rangle}(x) \leq 0 \text{ if } x \in \mathrm{supp}\mu(0) \ . \end{aligned}$$

(the generalized *Volterra averaging principle* [52]).

Formally, various definitions of genericity are constructed as follows. All systems (or cases, or situations and so on) under consideration are somehow parameterized – by sets of vectors, functions, matrices etc. Thus, the "space of systems" Q can be described. Then the *"thin sets"* are introduced into Q, i.e. the sets, which we shall later neglect. The union of a finite or countable number of thin sets, as well as the intersection of any number of them should be thin again, while the whole Q is not thin. There are two traditional ways to determine thinness.

1. A set is considered thin when it has *measure zero*. This is resonable for a finite-dimensional case, when there is the standard Lebesgue measure – the length, the area, the volume.
2. But most frequently we deal with the functional parameters. In that case it is common to restore to the second definition, according to which the sets of Baire first category are negligible. The construction begins with nowhere dense sets. The set Y is nowhere dense in Q, if in any nonempty open set $V \subset Q$ (for example, in a ball) there exists a nonempty open subset $W \subset V$ (for example, a ball), which does not intersect with Y: $W \cap Y = \emptyset$. Roughly speaking, Y is "full of holes" – in any neighborhood of any point of the set Y there is an open hole. Countable union of nowhere dense sets is called the set of first category. The second usual way is to define thin sets as the *sets of first category*. A *residual set* (a "thick" set) is the complement of a set of the first category.

For the second approach, the Baire category theorem is important: In a non-empty complete metric space, any countable intersection of dense, open subsets is non-empty.

But even the real line R can be divided into two sets, one of which has zero measure, the other is of first category. The genericity in the sense of measure and the genericity in the sense of category considerably differ in the applications where both of these concepts can be used. The conflict between the two main views on genericity stimulated efforts to invent new and stronger approaches.

Systems (1) were parameterized by continuous maps $\mu \mapsto k_\mu$. Denote by Q the space of these maps $M \to C(X)$ with the topology of uniform convergence on M. It is a Banach space. Therefore, we shall consider below thin sets in a Banach space Q. First of all, let us consider n-dimensional affine compact subsets of Q as a Banach space of affine maps $\Psi : [0,1]^n \to Q$ $(\Psi(\alpha_1, \dots \alpha_n) = \sum_i \alpha_i f_i + \varphi$, $\alpha_i \in [0,1]$, $f_i, \varphi \in Q)$ in the maximum norm. For the image of a map Ψ we use the standard notation $\mathrm{im}\Psi$.

Definition 2. A set $Y \subset Q$ is n-thin, if the set of affine maps $\Psi : [0,1]^n \to Q$ with non-empty intersection $\mathrm{im}\Psi \cap Y \neq \emptyset$ is the set of first category.

All compact sets in infinite-dimensional spaces and closed linear subspaces with codimension greater then n are n-thin. If $\dim Q \leq n$, then only empty set is n-thin in Q. The union of a finite or countable number of n-thin sets, as well as the intersection of any number of them is n-thin, while the whole Q is not n-thin.

Let us consider compact subsets in Q parametrized by points of a compact space K. It can be presented as a Banach space $C(K,Q)$ of continuous maps $K \to Q$ in the maximum norm.

Definition 3. A set $Y \subset Q$ is completely thin, if for any compact K the set of continuous maps $\Psi : K \to Q$ with non-empty intersection $\mathrm{im}\Psi \cap Y \neq \emptyset$ is the set of first category.

A set Y in the Banach space Q is completely thin, if for any compact set K in Q and arbitrary positive $\varepsilon > 0$ there exists a vector $q \in Q$, such that $\|q\| < \varepsilon$ and $K + q$ does not intersect Y: $(K + q) \cap Y = \emptyset$. All compact sets in infinite-dimensional spaces and closed linear subspaces with infinite codimension are completely thin. Only empty set is completely thin in a finite-dimensional space. The union of a finite or countable number of completely thin sets, as well as the intersection of any number of them is completely thin, while the whole Q is not completely thin.

Proposition 4. *If a set Y in the Banach space Q is completely thin, then for any compact metric space K the set of continuous maps $\Psi : K \to Q$ with non-empty intersection* $\mathrm{im}\Psi \cap Y \neq \emptyset$ *is completely thin in the Banach space $C(K, Q)$.* □

Below the wording "almost always" means: the set of exclusions is completely thin. The main result presented in this section sounds as follows: almost always the sets of global maxima of functions from a compact set are uniformly almost finite.

Proposition 5. *Let X have no isolated points. Then almost always a function $f \in C(X)$ has nowhere dense set of zeros* $\{x \in X \mid f(x) = 0\}$ *(the set of exclusions is completely thin in $C(X)$).* □

After combination Proposition 5 with Proposition 4 we get the following

Proposition 6. *Let X have no isolated points. Then for any compact space K and almost every continuous map $\Psi : K \to C(X)$ all functions $f \in \mathrm{im}\Psi$ have nowhere dense sets of zeros (the set of exclusions is completely thin in $C(K, C(X))$).* □

In other words, in almost every compact family of continuous functions all the functions have nowhere dense sets of zeros.

Let us consider a space of closed subsets of the compact metric space X endowed by the Hausdorff distance. The Hausdorff distance between closed subsets of X is

$$\mathrm{dist}(A, B) = \max\{\sup_{x \in A} \inf_{x \in B} \rho(x, y) \ , \ \sup_{x \in B} \inf_{x \in A} \rho(x, y)\} \ .$$

The almost finite sets were introduced in [21] for description of the typical sets of maxima for continuous functions from a compact set. This definition depends on an arbitrary sequence $\varepsilon_n > 0$, $\varepsilon_n \to 0$. For any such sequence we construct a class of subsets $Y \subset X$ that can be approximated by finite set faster than $\varepsilon_n \to 0$, and for families of sets we introduce a notion of *uniform* approximation by finite sets faster than $\varepsilon_n \to 0$:

Definition 7. Let $\varepsilon_n > 0$, $\varepsilon_n \to 0$. The set $Y \subset X$ can be approximated by finite sets faster than $\varepsilon_n \to 0$ ($\varepsilon_n > 0$), if for any $\delta > 0$ there exists a finite set S_N such that $\mathrm{dist}(S_N, Y) < \delta\varepsilon_N$. The sets of family $\mathbb{Y}$ can be uniformly approximated by finite sets faster than $\varepsilon_n \to 0$, if for any $\delta > 0$ there exists such a number N that for any $Y \in \mathbb{Y}$ there exists a finite set S_N such that $\mathrm{dist}(S_N, Y) < \delta\varepsilon_N$.

The simplest example of almost finite set on the real line for a given $\varepsilon_n \to 0$ ($\varepsilon_n > 0$) is the sequence ε_n/n. If $\varepsilon_n < \mathrm{const}/n$, then the set Y on the real line which can be approximated by finite sets faster than $\varepsilon_n \to 0$ have zero Lebesgue measure. At the same time, it is nowhere dense, because it can be covered by a finite number of intervals with an arbitrary small sum of lengths (hence, in any interval we can find a subinterval free of points of Y).

Let us study the sets of global maxima $\mathrm{argmax} f$ for continuous functions $f \in C(X)$. For each $f \in C(X)$ and any $\epsilon > 0$ there exists $\phi \in C(X)$ such that $\|f - \phi\| \leq \epsilon$ and $\mathrm{argmax}\phi$ consists of one point. Such a function ϕ can be chosen in the form

$$\phi(x) = f(x) + \frac{\epsilon}{1 + \rho(x, x_0)^2} ,$$

where x_0 is an arbitrary element of $\mathrm{argmax} f$. In this case $\mathrm{argmax}\phi = \{x_0\}$.

Hence, the set $\mathrm{argmax} f$ can be reduced to one point by an arbitrary small perturbations of the function f. On the other hand, it is impossible to extend significantly the set $\mathrm{argmax} f$ by a sufficiently small perturbation, the dependence of this set on f is semicontinuous in the following sense.

Proposition 8. *For given $f \in C(X)$ and any $\varepsilon > 0$ there exists $\delta > 0$ such that, whenever $\|f - \phi\| < \delta$, then*

$$\max_{x \in \mathrm{argmax}\phi} \; \min_{y \in \mathrm{argmax} f} \rho(x, y) < \varepsilon \ . \ \square \tag{15}$$

These constructions can be generalized onto n-parametric affine compact families of continuous functions. Let us consider affine maps of the cube $[0,1]^k$ into $C(X)$, $\varPhi : [0,1]^k \to C(X)$. The space of all such maps is a Banach space endowed with the maximum norm.

Proposition 9. *For any affine map $\varPhi : [0,1]^k \to C(X)$ and an arbitrary $\epsilon > 0$ there exists such a continuous function $\psi \in C(X)$, that $\|\psi\| < \epsilon$ and the set* $\mathrm{argmax}(f + \psi)$ *includes not more than $k+1$ points for all $f \in \mathrm{im}\varPhi$.*
$\square$

To prove this Proposition we used the following Lemma which is of general interest.

Lemma 10. *Let $Q \subset C(X)$ be a compact set of functions, $\varepsilon > 0$. Then there are a finite set $Y \subset X$ and a function $\phi \in C(X)$ such that $\|\phi\| < \varepsilon$, and any function $f \in Q + \phi$ achieves its maximum only on Y:* $\mathrm{argmax} f \subset Y$. $\square$

Note, that Proposition 9 and Lemma 10 demonstrate us different sources of discreteness: in Lemma 10 it is the approximation of a compact set by a finite net, and in Proposition 9 it is the connection between the number of parameters and the possible number of global maximums in a k-parametric family of functions. There is no direct connection between N and k values, and it might be that $N \gg k$. For smooth functions in finite-dimensional real space polynomial approximations can be used instead of Lemma 10 in order to prove the analogue of Proposition 9.

The rest of this Sec. 2.4 is devoted to application of Proposition 9 to evaluation of maximizers for functions from a compact sets of functions. For any compact K the space of continuous maps $C(K, C(X))$ is isomorphic to the space of continuous functions $C(K \times X)$. Each continuous map $F : K \to C(X)$ can be approximated with an arbitrary accuracy $\varepsilon > 0$ by finite sums of the following form ($k \geq 0$):

$$F(y)(x) = \sum_{i=1}^{k} \alpha_i(y) f_i(x) + \varphi(x) + o \ ,$$
$$y \in K,\ x \in X,\ 0 \leq \alpha_i \leq 1,\ f_i, \varphi \in C(X),\ |o| < \varepsilon \ . \tag{16}$$

Each set $f_i, \varphi \in C(X)$ generates a map $\Phi : [0,1]^k \to C(X)$. A dense subset in the space of these maps satisfy the statement of Proposition 9: each function from $\mathrm{im}\Phi$ has not more than $k+1$ points of global maximum. Let us use for this set of maps Φ notation $\mathbf{P}_k$, for the correspondent set of the maps $F : K \to C(X)$, which have the form of finite sums (16), notation $\mathbf{P}_k^K$, and $\mathbf{P}^K = \cup_k \mathbf{P}_k^K$.

For each $\Phi \in \mathbf{P}_k^K$ and any $\varepsilon > 0$ there is $\delta = \delta_\Phi(\varepsilon) > 0$ such that, whenever $\|\Psi - \Phi\| < \delta_\Phi(\varepsilon)$, the set $\mathrm{argmax} f$ belongs to a union of $k+1$ balls of radius ε for any $f \in \mathrm{im}\Psi$ (Proposition 8).

Let us introduce some notations: for $k \geq 0$ and $\varepsilon > 0$

$$\mathbf{U}_{k,\varepsilon}^K = \{\Psi \in C(K, C(X)) \,|\, \|\Psi - \Phi\| < \delta_\Phi(\varepsilon) \text{ for some } \Phi \in \mathbf{P}_k^K\} \ ;$$

for $\varepsilon_i > 0$, $\varepsilon_i \to 0$

$$\mathbf{V}_{\{\varepsilon_i\}}^K = \bigcup_{k=0}^{\infty} \mathbf{U}_{k,\varepsilon_k}^K \ ;$$

and, finally,

$$\mathbf{W}_{\{\varepsilon_i\}}^K = \bigcap_{s=1}^{\infty} \mathbf{V}_{\{\frac{1}{2^s}\varepsilon_i\}}^K \ .$$

The set $\mathbf{P}^K$ is dense in $C(K, C(X))$. Any $F \in \mathbf{P}^K$ has the form of finite sum (16), and any $f \in \mathrm{im}F$ has not more than $k+1$ point of global maximum, where k is the number of summands in presentation (16). The sets $\mathbf{V}^K_{\{\varepsilon_i\}}$ are open and dense in the Banach space $C(K, C(X))$ for any sequence $\varepsilon_i > 0$, $\varepsilon_i \to 0$. The set $\mathbf{W}^K_{\{\varepsilon_i\}}$ is intersection of countable number of open dense sets. For any $F \in \mathbf{W}^K_{\{\varepsilon_i\}}$ the sets of the family $\{\mathrm{argmax} f \mid f \in \mathrm{im}F\}$ can be uniformly approximated by finite sets faster than $\varepsilon_n \to 0$. It is proven that this property is typical in the Banach space $C(K, C(X))$ in the sense of category.

In order to prove that the set of exclusions is completely thin in $C(K, C(X))$ it is sufficient to use the approach of Proposition 4. Note that for arbitrary compact space Q the set of continuous maps $Q \to C(K, C(X))$ in the maximum norm is isomorphic to the spaces $C(Q \times K, C(X))$ and $C(Q \times K \times X)$. The space $Q \times K$ is compact. We can apply the previous construction to the space $C(Q \times K, C(X))$ for arbitrary compact Q and get the result: the set of exclusion is completely thin in $C(K, C(X))$.

In the definition of $\mathbf{W}^K_{\{\varepsilon_i\}}$ we use only one sequence $\varepsilon_i > 0$, $\varepsilon_i \to 0$. Of course, for any finite or countable set of sequences the intersection of correspondent sets $\mathbf{W}^K_{\{\varepsilon_i\}}$ is also a residual set, and we can claim that almost always the sets of $\{\mathrm{argmax} f \mid f \in \mathrm{im}F\}$ can be uniformly approximated by finite sets faster than $\varepsilon_n \to 0$ for all given sequences.

2.5 Selection Efficiency

The first application of the extremal principle for the ω-limit sets is the theorem of the selection efficiency. The dynamics of a system with inheritance leads indeed to a selection in the limit $t \to \infty$. In the typical situation, a diversity in the limit $t \to \infty$ becomes less than the initial diversity. There is an efficient selection for the "best". The basic effects of selection are formulated below. Let X be compact metric space without isolated points.

Theorem 11. (Theorem of selection efficiency.)

1. *For almost every system (1) the support of any ω-limit distribution is nowhere dense in X (and it has the Lebesgue measure zero for Euclidean space).*
2. *Let $\varepsilon_n > 0$, $\varepsilon_n \to 0$ be an arbitrary chosen sequence. The following statement is true for almost every system (1). Let the support of the initial distribution be the whole X. Then the support of any ω-limit distribution can be approximated by finite sets uniformly faster than $\varepsilon_n \to 0$.*

The set of exclusive systems that do not satisfy the statement 1 or 2 is completely thin.

Remark. These properties hold for the continuous reproduction coefficients. It is well-known, that it is dangerous to rely on the genericity among continuous functions. For example, almost all continuous functions are nowhere differentiable. But the properties 1, 2 hold also for the smooth reproduction coefficients on the manifolds and sometimes allow to replace the "almost finiteness" by simply finiteness.

Scheme of Proof. To prove the first statement, it is sufficient to refer to Proposition 2.3. In order to clarify the second part of this theorem, note that:

1. Support of an arbitrary ω-limit distribution μ^* consist of points of global maximum of the average reproduction coefficient on a support of the initial distribution. The corresponding maximum value is zero.
2. Almost always a function has only one point of global maximum, and corresponding maximum value is not 0.
3. In a one-parametric family of functions almost always there may occur zero values of the global maximum (at one point), which cannot be eliminated by a small perturbation, and individual functions of the family may stably have two global maximum points.
4. For a generic n-parameter family of functions, there may exist stably a function with n points of global maximum and with zero value of this maximum.
5. Our phase space M is compact. The set of corresponding reproduction coefficients k_M in $C(X)$ for the given map $\mu \to k_\mu$ is compact too. The average reproduction coefficients belong to the closed convex hull of this set $\overline{\text{conv}}(k_M)$. And it is compact too.
6. A compact set in a Banach space can be approximated by compacts from finite-dimensional linear manifolds. Generically, in a space of continuous functions, a function, which belongs such a n-dimensional compact, can have not more than n points of global maximum with zero maximal value.

The rest of the of proof of the second statement is purely technical. Some technical details are presented in the previous section. The easiest demonstration of the "natural" character of these properties is the demonstration of instability of exclusions: If, for example, a function has several points of global maxima, then with an arbitrary small perturbation (for all usually used norms) it can be transformed into a function with the unique point of global maximum. However "stable" does not always mean "dense". The discussed properties of the system (1) are valid in a very strong sense: the set of exclusion is completely thin. □

2.6 Gromov's Interpretation of Selection Theorems

In his talk [26], M. Gromov offered a geometric interpretation of the selection theorems. Let us consider dynamical systems in the standard m-simplex σ_m

in $m+1$-dimensional space R^{m+1}:

$$\sigma_m = \{x \in R^{m+1} \,|\, x_i \geq 0, \ \sum_{i=1}^{m+1} x_i = 1\} \ .$$

We assume that simplex σ_m is positively invariant with respect to these dynamical systems: if the motion starts in σ_m at some time t_0, then it remains in σ_m for $t > t_0$. Let us consider the motions that start in the simplex σ_m at $t = 0$ and are defined for $t > 0$.

For large m, almost all volume of the simplex σ_m is concentrated in a small neighborhood of the center of σ_m, near the point $c = \left(\frac{1}{m}, \frac{1}{m}, \dots, \frac{1}{m}\right)$. Hence, one can expect that a typical motion of a general dynamical system in σ_m for sufficiently large m spends almost all the time in a small neighborhood of c.

Indeed, the m-dimensional volume of σ_m is $V_m = \frac{1}{m!}$. The part of σ_m, where $x_i \geq \varepsilon$, has the volume $(1-\varepsilon)^m V_m$. Hence, the part of σ_m, where $x_i < \varepsilon$ for all $i = 1, \dots, m+1$, has the volume $V_\varepsilon > (1-(m+1)(1-\varepsilon)^m)V_m$. Note, that $(m+1)(1-\varepsilon)^m \sim m\exp(-\varepsilon m) \to 0$, if $m \to \infty$ $(1 > \varepsilon > 0)$. Therefore, for $m \to \infty$, $V_\varepsilon = (1-o(1))V_m$. The volume W_ρ of the part of σ_m with Euclidean distance to the center c less than $\rho > 0$ can be estimated as follows: $W_\rho > V_\varepsilon$ for $\varepsilon\sqrt{m+1} = \rho$, hence $W_\rho > (1-(m+1)(1-\rho/\sqrt{m+1})^m)V_m$. Finally, $(m+1)(1-\rho/\sqrt{m+1})^m \sim m\exp(-\rho\sqrt{m})$, and $W_\rho = (1-o(1))V_m$ for $m \to \infty$. Let us mention here the opposite concentration effect for a m-dimensional ball B_m: for $m \to \infty$ the most part of its volume is concentrated in an arbitrary small vicinity of its boundary, the sphere. This effect is the essence of the famous equivalence of microcanonical and canonical ensembles in statistical physics (for detailed discussion see [22]).

Let us consider dynamical systems with an additional property ("inheritance"): all the faces of the simplex σ_m are also positively invariant with respect to the systems with inheritance. It means that if some $x_i = 0$ initially at the time $t = 0$, then $x_i = 0$ for $t > 0$ for all motions in σ_m. The essence of selection theorems is as follows: a typical motion of a typical dynamical system with inheritance spends almost all the time in a small neighborhood of low-dimensional faces, even if it starts near the center of the simplex.

Let us denote by $\partial_r\sigma_m$ the union of all r-dimensional faces of σ_m. Due to the selection theorems, a typical motion of a typical dynamical system with inheritance spends almost all time in a small neighborhood of $\partial_r\sigma_m$ with $r \ll m$. It should not obligatory reside near just one face from $\partial_r\sigma_m$, but can travel in neighborhood of different faces from $\partial_r\sigma_m$ (the drift effect). The minimax estimation of the number of points in ω-limit distributions through the diameters $\varepsilon_n > 0$ of the set $\overline{\mathrm{conv}}(k_M)$ is the estimation of r.

3 Drift and Mutations

3.1 Drift Equations

So far, we talked about the support of an individual ω-limit distribution. For almost all systems it is small. But this does not mean, that the union of these supports is small even for one solution $\mu(t)$. It is possible that a solution is a finite set of narrow peaks getting in time more and more narrow, moving slower and slower, but not tending to fixed positions, rather continuing to move along its trajectory, and the path covered tends to infinity as $t \to \infty$.

This effect was not discovered for a long time because the slowing down of the peaks was thought as their tendency to fixed positions For the best of our knowledge, the first detailed publication of the drift equations and corresponded types of stability appeared in book [21], first examples of coevolution drift on a line were published in the series of papers [46].

There are other difficulties related to the typical properties of continuous functions, which are not typical for the smooth ones. Let us illustrate them for the distributions over a straight line segment. Add to the reproduction coefficients k_μ the sum of small and narrow peaks located on a straight line distant from each other much more than the peak width (although it is ε-small). However small is chosen the peak's height, one can choose their width and frequency on the straight line in such a way that from any initial distribution μ_0 whose support is the whole segment, at $t \to \infty$ we obtain ω-limit distributions, concentrated at the points of maximum of the added peaks.

Such a model perturbation is small in the space of continuous functions. Therefore, it can be put as follows: *by small continuous perturbation the limit behavior of system (1) can be reduced onto a ε-net for sufficiently small ε*. But this can not be done with the small smooth perturbations (with small values of the first and the second derivatives) in the general case. The discreteness of the net, onto which the limit behavior is reduced by small continuous perturbations, differs from the discreteness of the support of the individual ω-limit distribution. For an individual distribution the number of points is estimated, roughly speaking, by the number of essential parameters (11), while for the conjunction of limit supports – by the number of stages in approximation of k_μ by piece-wise constant functions.

Thus, in a typical case the dynamics of systems (1) with smooth reproduction coefficients transforms a smooth initial distributions into the ensemble of narrow peaks. The peaks become more narrow, their motion slows down, but not always they tend to fixed positions.

The equations of motion for these peaks can be obtained in the following way [21]. Let X be a domain in the n-dimensional real space, and the initial distributions μ_0 be assumed to have smooth density. Then, after sufficiently large time t, the position of distribution peaks are the points of the average reproduction coefficient maximum $\langle k_\mu \rangle_t$ (9) to any accuracy set in advance.

Let these points of maximum be x^α, and

$$q_{ij}^\alpha = -t\frac{\partial^2 \langle k_\mu \rangle_t}{\partial x_i \partial x_j}\Big|_{x=x^\alpha} .$$

It is easy to derive the following differential relations just by differentiation in time of the extremum conditions: at points $x^\alpha(t)$ gradient of the average reproduction coefficient $\langle k_\mu \rangle_t$ vanishes: $\partial \langle k_\mu \rangle_t(x)/\partial x_i \,|_{x=x^\alpha(t)} = 0$.

$$\begin{aligned} \sum_j q_{ij}^\alpha \frac{\mathrm{d}x_j^\alpha}{\mathrm{d}t} &= \frac{\partial k_{\mu(t)}}{\partial x_i}\Big|_{x=x^\alpha} ; \\ \frac{\mathrm{d}q_{ij}^\alpha}{\mathrm{d}t} &= -\frac{\partial^2 k_{\mu(t)}}{\partial x_i \partial x_j}\Big|_{x=x^\alpha} . \end{aligned} \tag{17}$$

The exponent coefficients q_{ij}^α remain time dependent even when the distribution tends to a δ-function. It means (in this case) that peaks became infinitely narrow. Nevertheless, it is possible to change variables and represent the weak* tendency to stationary discrete distribution as usual tendency to a fixed points, see (21) below.

These relations (17) do not form a closed system of equations, because the right-hand parts are not functions of x_i^α and q_{ij}^α. For sufficiently narrow peaks there should be separation of the relaxation times between the dynamics *on* the support and the dynamics *of* the support: the relaxation of peak amplitudes (it can be approximated by the relaxation of the distribution with the finite support, $\{x^\alpha\}$) should be significantly faster than the motion of the locations of the peaks, the dynamics of $\{x^\alpha\}$. Let us write the first term of the corresponding asymptotics [21].

For the finite support $\{x^\alpha\}$ the distribution is $\mu = \sum_\alpha N_\alpha \delta(x - x^\alpha)$. Dynamics of the finite number of variables, N_α obeys the system of ordinary differential equations

$$\frac{\mathrm{d}N_\alpha}{\mathrm{d}t} = k_\alpha(N) N_\alpha, \tag{18}$$

where N is vector with components N_α, $k_\alpha(N)$ is the value of the reproduction coefficient k_μ at the point x^α:

$$k_\alpha(N) = k_\mu(x^\alpha) \text{ for } \mu = \sum_\alpha N_\alpha \delta(x - x^\alpha) .$$

For finite-dimensional dynamics (18) we have to find the relevant SBR (Sinai–Bowen–Ruelle) invariant measure (or "physical measure") [30,34] for averaging and substitute the average time along the solutions of (18)

$$\frac{1}{t}\int_0^t k_{\mu^*(N)(\tau)}\,\mathrm{d}\tau \quad \text{where} \quad \mu^*(N) = \sum_\alpha N_\alpha \delta(x - x^\alpha)$$

by the average with respect to the SBR measure on space of vectors N. For this average, we use notation $k^*(\{x^\alpha\}) = \langle k_{\mu^*} \rangle$.

In the simplest case the finite-dimensional attractor is just one stable fixed point and the average $k^*(\{x^\alpha\})$ is a value at this point. Let the dynamics of the system (18) for a given set of initial conditions be simple: the motion $N(t)$ goes to the stable fixed point $N = N^*(\{x^\alpha\})$. Then we can take $k^*(\{x^\alpha\}) = k_{\mu^*}$ where $\mu^* = \sum_\alpha N^*_\alpha \delta(x - x^\alpha)$.

One can use in the right hand side of (17) the following approximation for $k^*(\{x^\alpha\})$ instead of $k_{\mu(t)}$.

$$\mu(t) = \mu^*(\{x^\alpha(t)\}) = \sum_\alpha N^*_\alpha \delta(x - x^\alpha(t)) \ . \tag{19}$$

This is a standard averaging hypothesis. We can use it because density peaks are sufficiently narrow, hence, (i) the difference between true $k_{\mu(t)}$ and the reproduction coefficient for the measure with finite support $k(\sum_\alpha N_\alpha(t)\delta(x - x^\alpha$ is negligible and (ii) dynamics of peak motion is much slower than relaxation of the finite-dimensional system (18) to its attractor.The relations (17) transform into the ordinary differential equations

$$\sum_j q^\alpha_{ij} \frac{\mathrm{d}x^\alpha_j}{\mathrm{d}t} = \left.\frac{\partial k^*(\{x^\beta\})(x)}{\partial x_i}\right|_{x=x^\alpha} \ ;$$
$$\frac{\mathrm{d}q^\alpha_{ij}}{\mathrm{d}t} = -\left.\frac{\partial^2 k^*(\{x^\beta\})(x)}{\partial x_i \partial x_j}\right|_{x=x^\alpha} \ . \tag{20}$$

The matrix variables q^α_{ij} are usually not bounded. For example, near a non-degenerated fixed point $\{x^\alpha\}$ they go to infinity linearly in time. On the other hand, relaxation of $\{x^\alpha\}$ to their stationary positions, for example, is not exponential due to (20). To return to the standard situation with compact phase space and exponential relaxation it is useful to switch to the logarithmic time $\tau = \ln t$ and to new variables

$$b^\alpha_{ij} = \frac{1}{t} q^\alpha_{ij} = -\left.\frac{\partial^2 \langle k(\mu)\rangle_t}{\partial x_i \partial x_j}\right|_{x=x^\alpha} \ .$$

For large t we obtain from (20)

$$\sum_j b^\alpha_{ij} \frac{\mathrm{d}x^\alpha_j}{\mathrm{d}\tau} = \left.\frac{\partial k^*(\{x^\beta\})(x)}{\partial x_i}\right|_{x=x^\alpha} \ ;$$
$$\frac{\mathrm{d}b^\alpha_{ij}}{\mathrm{d}\tau} = -\left.\frac{\partial^2 k^*(\{x^\alpha\})(x)}{\partial x_i \partial x_j}\right|_{x=x^\beta} - b^\alpha_{ij} \ . \tag{21}$$

In these equations it becomes obvious that dynamics of matrix b^α_{ij} is the differential pursuit of Hessian $\partial^2 k^*(\{x^\alpha\})(x)/\partial x_i \partial x_j \,|_{x=x^\beta}$.

Equations for drift in logarithmic time (21) are the main equations in the theory of the asymptotic layer For Darwin's systems near their limit behavior.

The way of constructing the drift equations (20,21) for a specific system (1) is as follows:

1. For finite sets $\{x^\alpha\}$ one studies systems (18) and finds the equilibrium solutions $^*(\{x^\alpha\})$ or the relevant SBR measure;
2. For given measures $\{x^\alpha\}$ (19) one calculates the reproduction coefficients $k^*(\{x^\alpha\})(x)$ together with the first and second first derivatives of these functions in x at points x^α. That is all, the drift equations (21) are set up.

The drift equations (20,21) describe the dynamics of the peaks positions x^α and of the coefficients q_{ij}^α. For given x^α, q_{ij}^α and N_α^* the distribution density μ can be approximated as the sum of narrow Gaussian peaks:

$$\mu = \sum_\alpha N_\alpha^* \sqrt{\frac{\det Q^\alpha}{(2\pi)^n}} \exp\left(-\frac{1}{2}\sum_{ij} q_{ij}^\alpha (x_i - x_i^\alpha)(x_j - x_j^\alpha)\right) \ , \qquad (22)$$

where Q^α is the inverse covariance matrix (q_{ij}^α).

If the limit dynamics of the system (18) for finite supports at $t \to \infty$ can be described by a more complicated attractor, then instead of reproduction coefficient $k^*(\{x^\alpha\})(x) = k_{\mu^*}$ for the stationary measures μ^* (19) one can use the average reproduction coefficient with respect to the corresponding Sinai–Ruelle–Bowen measure. If finite systems (18) have several attractors for given $\{x^\alpha\}$, then the dependence $k^*(\{x^\alpha\})$ is multi-valued, and there may be bifurcations and hysteresis with the function $k^*(\{x^\alpha\})$ transition from one sheet to another. There are many interesting effects concerning peaks' birth, desintegration, divergence, and death, and the drift equations (20,21) describe the motion in a non-critical domain, between these critical effects.

Inheritance (conservation of support) is never absolutely exact. Small variations, mutations, immigration in biological systems are very important. Excitation of new degrees of freedom, modes diffusion, noise are present in physical systems. How does small perturbation in the inheritance affect the effects of selection? The answer is usually as follows: there is such a value of perturbation of the right-hand side of (1), at which they would change nearly nothing, just the limit δ-shaped peaks transform into sufficiently narrow peaks, and zero limit of the velocity of their drift at $t \to \infty$ substitutes by a small finite one.

3.2 Drift in Presence of Mutations and Scaling Invariance

The simplest model for "inheritance + small variability" is given by a perturbation of (1) with diffusion term

$$\frac{\partial \mu(x,t)}{\partial t} = k_{\mu(x,t)} \times \mu(x,t) + \varepsilon \sum_{ij} d_{ij}(x) \frac{\partial^2 \mu(x,t)}{\partial x_i \partial x_j} \ . \qquad (23)$$

where $\varepsilon > 0$ and the matrix of diffusion coefficients d_{ij} is symmetric and positively definite.

There are almost always no qualitative changes in the asymptotic behavior, if ε is sufficiently small. With this the asymptotics is again described by the drift equations (20,21), modified by taking into account the diffusion as follows:

$$\sum_j q_{ij}^{\alpha} \frac{\mathrm{d}x_j^{\alpha}}{\mathrm{d}t} = \left. \frac{\partial k^*(\{x^{\beta}\})(x)}{\partial x_i} \right|_{x=x^{\alpha}} ;$$

$$\frac{\mathrm{d}q_{ij}^{\alpha}}{\mathrm{d}t} = - \left. \frac{\partial^2 k^*(\{x^{\beta}\})(x)}{\partial x_i \partial x_j} \right|_{x=x^{\alpha}} - 2\varepsilon \sum_{kl} q_{ik}^{\alpha} d_{kl}(x^{\alpha}) q_{lj}^{\alpha} . \qquad (24)$$

Now, as distinct from (20), the eigenvalues of the matrices $Q^{\alpha} = (q_{ij}^{\alpha})$ cannot grow infinitely. This is prevented by the quadratic terms in the right-hand side of the second equation (24).

Dynamics of (24) does not depend on the value $\varepsilon > 0$ qualitatively, because of the obvious scaling property. If ε is multiplied by a positive number ν, then, upon rescalling $t' = \nu^{-1/2} t$ and $q_{ij}^{\alpha}{}' = \nu^{-1/2} q_{ij}^{\alpha}$, we have the same system again. Multiplying $\varepsilon > 0$ by $\nu > 0$ changes only peak's velocity values by a factor $\nu^{1/2}$, and their width by a factor $\nu^{1/4}$. The paths of peaks' motion do not change at this for the drift approximation (24) (but the applicability of this approximation may, of course, change).

4 Three Main Types of Stability

4.1 Internal stability

Stable steady-state solutions of equations of the form (1) may be only the sums of δ-functions – this was already mentioned. There is a set of specific conditions of stability, determined by the form of equations.

Consider a stationary distribution for (1) with a finite support

$$\mu^*(x) = \sum_{\alpha} N_{\alpha}^* \delta(x - x^{*\alpha}) .$$

Steady state of μ^* means, that

$$k_{\mu^*}(x^{*\alpha}) = 0 \text{ for all } \alpha . \qquad (25)$$

The *internal stability* means, that this distribution is stable with respect to perturbations not increasing the support of μ^*. That is, the vector N_{α}^* is the stable fixed point for the dynamical system (18). Here, as usual, it is possible to distinguish between the Lyapunov stability, the asymptotic stability and the first approximation stability (negativeness of real parts for the eigenvalues of the matrix $\partial \dot{N}_{\alpha}^* / \partial N_{\alpha}^*$ at the stationary points).

4.2 External Stability – Uninvadability

The *external stability* (*uninvadability*) means stability to an expansion of the support, i.e. to adding to μ^* of a small distribution whose support contains points not belonging to suppμ^*. It makes sense to speak about the external stability only if there is internal stability. In this case it is sufficient to restrict ourselves with δ-functional perturbations. The external stability has a very transparent physical and biological sense. It is stability with respect to *introduction* into the systems of a new inherited unit (gene, variety, specie...) in a small amount.

The *necessary condition for the external stability* is: the points $\{x^{*\alpha}\}$ are points of the global maximum of the reproduction coefficient $k_{\mu^*}(x)$. It can be formulated as the optimality principle

$$k_{\mu^*}(x) \leq 0 \text{ for all } x;\; k_{\mu^*}(x^{*\alpha}) = 0 \ . \tag{26}$$

The *sufficient condition for the external stability* is: the points $\{x^{*\alpha}\}$ and only these points are points of the global maximum of the reproduction coefficient $k_{\mu^*}(x^{*\alpha})$. At the same time it is the condition of the external stability in the first approximation and the optimality principle

$$k_{\mu^*}(x) < 0 \text{ for } x \notin \{x^{*\alpha}\};\; k_{\mu^*}(x^{*\alpha}) = 0 \ . \tag{27}$$

The only difference from (26) is the change of the inequality sign from $k_{\mu^*}(x) \leq 0$ to $k_{\mu^*}(x) < 0$ for $x \notin \{x^{*\alpha}\}$. The necessary condition (26) means, that the small δ-functional addition will not grow in the first approximation. According to the sufficient condition (27) such a small addition will exponentially decrease.

If X is a finite set, then the combination of the external and the internal stability is equivalent to the standard stability for a system of ordinary differential equations.

4.3 Stable Realizability – Evolutionary Robustness

External stability of a internally stable limit distribution is insufficient for its stability with respect to the drift: It does not imply convergence to x^* when starting from a distribution of small deviations from x^*, regardless of how small these deviations are. The standard idea of asymptotic stability is: "after small deviation the system returns to the initial regime, and do not deviate to much on the way of returning". The crucial question for the measure dynamics is: in which topology the deviation is small? The small shift of the narrow peak of distribution in the continuous space of strategies can be considered as a small deviation in the weak* topology, but it is definitely large deviation in the strong topology, for example, if the shift is not small in comparison with the peak with.

For the continuous X there is one more kind of stability important from the applications viewpoint. Substitute δ-shaped peaks at the points $\{x^{*\alpha}\}$ by narrow Gaussians and shift slightly the positions of their maxima away from the points $x^{*\alpha}$. How will the distribution from such initial conditions evolve? If it tends to μ without getting too distant from this steady state distribution, then we can say that the third type of stability – *stable realizability* – takes place. It is worth mentioning that the perturbation of this type is only weakly* small, in contrast to perturbations considered in the theory of internal and external stability. Those perturbations are small by their norms. Let us remind that the norm of the measure μ is $\|\mu\| = \sup_{|f|\leq 1}[\mu, f]$. If one shifts the δ-measure of unite mass by any nonzero distance ε, then the norm of the perturbation is 2. Nevertheless, this perturbation weakly* tends to 0 with $\varepsilon \to 0$.

In order to formalize the condition of stable realizability it is convenient to use the drift equations in the form (21). Let the distribution μ^* be internally and externally stable in the first approximations. Let the points $x^{*\alpha}$ of global maxima of $k_{\mu^*}(x)$ be non-degenerate in the second approximation. This means that the matrices

$$b_{ij}^{*\alpha} = -\left(\frac{\partial^2 k_{\mu^*}(x)}{\partial x_i \partial x_j}\right)_{x=x^{*\alpha}} \tag{28}$$

are strictly positively definite for all α.

Under these conditions of stability and non-degeneracy the coefficients of (21) can be easily calculated using Taylor series expansion in powers of $(x^\alpha - x^{*\alpha})$. The stable realizability of μ^* in the first approximation means that the fixed point of the drift equations (21) with the coordinates

$$x^\alpha = x^{*\alpha}\ ,\ b_{ij}^{\alpha} = b_{ij}^{*\alpha} \tag{29}$$

is stable in the first approximation. It is the usual stability for the system (21) of ordinary differential equations, and these conditions with the notion of the stable realizability became clear from the logarithmic time drift equations (21) directly.

The specific structure of equations (21) allows us to simplify stability analysis for steady states. Let the steady state be externally stable steady states in the first approximations and let matrices $b^{*\alpha}$ be strictly positive definite. Equations (21) have the structure

$$\begin{aligned} \mathcal{B}\dot{\mathcal{X}} &= F(\mathcal{X})\,; \\ \dot{\mathcal{B}} &= \Phi(\mathcal{X}) - \mathcal{B}\,, \end{aligned} \tag{30}$$

where $\mathcal{X}$ is a vector composed from vectors x^α, and $\mathcal{B}$ is a block-diagonal matrix composed from matrices b^α. The steady values are $\mathcal{X}^*$ and $\mathcal{B}^*$: $F(\mathcal{X}^*) = 0$, $\mathcal{B}^* = \Phi(\mathcal{X}^*)$. Direct calculations gives for Jacobian J:

$$J = \begin{pmatrix} \mathcal{B}^{-1}\left.\frac{DF(\mathcal{X})}{D\mathcal{X}}\right|_{\mathcal{X}^*} & 0 \\ \left.\frac{D\Phi(\mathcal{X})}{D\mathcal{X}}\right|_{\mathcal{X}^*} & -1 \end{pmatrix} . \tag{31}$$

This form of Jacobian immedially implies the following proposition.

Proposition 12. *Stability of (21) (in the first approximation) near a steady state could be defined by the spectrum of matrix* $\mathcal{B}^{-1}\frac{DF(\mathcal{X})}{D\mathcal{X}}\Big|_{\mathcal{X}^*}$*: If the real parts of its eigenvalues are negative then the system is stable. If some of them are positive then the system is unstable.* □

To explain the sense of the stable realizability we used in the book [25] the idea of the "Gardens of Eden" from J. H. Conway "Game of Life" [18]. That are Game of Life patterns which have no father patterns and therefore can occur only at generation 0, from the very beginning. It is not known if a pattern which has a father pattern, but no grandfather pattern exists. It is the same situation, as for internal and external stable (uninvadable) state which is not stable realizable: it cannot be destroyed by mutants invasion and by the small variation of conditions, but, at the same time, it is not attractive for drift, and, hence, can not be realized in this asymptotic motion. It can be only created.

5 Explicit Drift Equations for Distributed Lottka–Volterra–Gause Systems

Construction of the drift system (21) goes through several operations. The most complicated of them is averaging: for a system with finite support $\{x^\alpha\}$ (18) we have to find the relevant finite-dimensional average $k^*(\{x^\alpha\}) = \langle k(\sum_\alpha N_\alpha \delta(x-x^\alpha))\rangle$

The most difficult operation in the construction of drift equations is the qualitative study of the finite-dimensional system and its SBR measures. Even in the simplest case of unique and globally attractive stable steady state in (18) the study of global stability may be difficult and even solution of equations for steady states may be computationally expensive.

There is a lucky exclusion: if the reproduction coefficient is a value of a linear integral operator then the steady-state can be found from linear equations and average values of k coincide with its values at steady-states. Replicator systems with linear reproduction coefficients include all classical Lottka–Volterra–Gause systems and can have an arbitrary complex dynamics. Nevertheless, equilibria of these systems satisfy linear equations and average reproduction coefficient is equal to its steady-state value.

Let us write down these equations:

$$\frac{\mathrm{d}\mu(x)}{\mathrm{d}t} = \mu(x)\left[r(x) + \int_X K(x,\xi)\mu(\xi)\,\mathrm{d}\xi\right] . \tag{32}$$

Here we assume that $q(x)$ and $K(x,\xi)$ are continuous functions.'The space of measures is assumed to have a bounded set of positive measures M $\mu(x \geq 0$ that is positively invariant relative to system (32): if $\mu(0) \in M$, then $\mu(t) \in M$

(we also assume that M is non-trivial, i.e. it is neither empty nor a one-point set but includes at least one point with its vicinity). This M serves as the phase space of system (32).

A steady state μ with support suppμ satisfies a linear equation:

$$r(x) + \int_X K(x,\xi)\mu(\xi)\,\mathrm{d}\xi = 0 \quad \text{for} \quad x \in \mathrm{supp}\mu \ . \tag{33}$$

For a finite set $\{x^\alpha\}$ we introduce a matrix $K(\{x^\alpha\}) = [K_{\alpha\beta}(\{x^\alpha\})] = [K(x^\alpha, x^\beta)]$ and a vector $r = r_\alpha = r(x^\alpha)$. Equation (18) for (32) has a form:

$$\frac{\mathrm{d}N_\alpha}{\mathrm{d}t} = \left(r_\alpha + \sum_\beta K_{\alpha\beta} N_\beta \right) N_\alpha, \tag{34}$$

The positive stationary solution (if it exists) is given by

$$N^*(\{x^\alpha\}) = -K^{-1}(\{x^\alpha\}) r(\{x^\alpha\}) \ . \tag{35}$$

For strictly positive bounded solutions of (34) $N(t)$ with $N(t) > \varepsilon > 0$ the time average of $N(t)$ coincides with N^* and the time average of any linear functional $l(N(t))$ is $l(N^*)$. Hence, for this type of finite-dimensional dynamics we can use in (21)

$$k^*(\{x^\alpha\})(x) = r(x) + \sum_\alpha K(x, x^\beta) N^*_\beta(\{x^\alpha\}) \ . \tag{36}$$

Here functions $N^*_\beta(\{x^\alpha\})$ are explicitly derived from the coefficients (35), hence. the drift equations (21) could be also found in explicit form, by differentiation. The coefficients of those equations includes nothing more than functions $r(x)$, $K(x, x^\beta)$, their rational combinations and derivatives.

Just for simplicity let us demonstrate this for system of two quasispecies. Let X be a disjoint union of two intervals on real line. The replicator system (32) in this case is

$$\begin{aligned}
\frac{\mathrm{d}\mu_1(x)}{\mathrm{d}t} &= \mu_1(x) \left[r_1(x) + \int K_{11}(x,\xi)\mu_1(\xi)\,\mathrm{d}\xi + \int K_{12}(x,\nu)\mu_2(\nu)\,\mathrm{d}\nu \right] ; \\
\frac{\mathrm{d}\mu_2(y)}{\mathrm{d}t} &= \mu_2(y) \left[r_2(x) + \int K_{21}(y,\xi)\mu_1(\xi)\,\mathrm{d}\xi + \int K_{22}(y,\nu)\mu_2(\nu)\,\mathrm{d}\nu \right] .
\end{aligned} \tag{37}$$

Two quasispecies are two peaks, one for μ_1 with coordinate $x = x^1$ and variance $\mathrm{var}_1 \approx 1/(tb^1)$ and another for μ_2 with coordinate $y = x^2$ and variance $\mathrm{var}_2 \approx 1/(tb^2)$ for large t.

The finite-dimensional system (34) transforms in

$$\begin{aligned}
\dot{N}_1 &= [r^1(x^1) + K_{11}(x^1, x^1)N_1 + K_{12}(x^1, x^2)N_2]N_1 \ ; \\
\dot{N}_2 &= [r^2(x^2) + K_{21}(x^2, x^1)N_1 + K_{22}(x^2, x^2)N_2]N_2 \ .
\end{aligned} \tag{38}$$

The steady state solution is

$$\begin{aligned} N_1^*(x^1,x^2) &= \frac{K_{22}r_1 - K_{12}r_2}{K_{12}K_{21} - K_{11}K_{22}}\,; \\ N_2^*(x^1,x^2) &= \frac{K_{11}r_2 - K_{21}r_1}{K_{12}K_{21} - K_{11}K_{22}}\,, \end{aligned} \tag{39}$$

where coefficients on the right hand side are calculated for $x = x^1$, $y = x^2$. The inequalities $N_{1,2}^* > 0$ should hold (this is a condition on the coefficients values). We omit here stability and positivity analysis for 2D system (38).

Function $k^*(\{x^\alpha\})(x)$ is represented by two functions $k_1^*(x^1,x^2)(x)$, $k_2^*(x^1,x^2)(y)$ (because X is a disjoint union of two intervals on real line):

$$\begin{aligned} k_1^*(x^1,x^2)(x) &= r_1(x) + K_{11}(x,x^1)N_1^*(x^1,x^2) + K_{12}(x,x^2)N_2^*(x^1,x^2)\,; \\ k_1^*(x^1,x^2)(y) &= r_2(y) + K_{21}(y,x^1)N_1^*(x^1,x^2) + K_{22}(y,x^2)N_2^*(x^1,x^2)\,. \end{aligned} \tag{40}$$

Now, we can write down the drift equation in logarithmic time (21):

$$\begin{aligned} \dot{x^1} &= \frac{1}{b^1}(\partial_x k_1^*(x^1,x^2)(x))_{x=x^1}\,, \quad \dot{x^2} = \frac{1}{b^2}(\partial_y k_2^*(x^1,x^2)(y))_{y=x^2}\,; \\ \dot{b^1} &= -(\partial_x^2 k_1^*(x^1,x^2)(x))_{x=x^1} - b^1\,, \;\; \dot{b^2} = -(\partial_y^2 k_2^*(x^1,x^2)(y))_{y=x^2} - b^2\,. \end{aligned} \tag{41}$$

Dynamics of $b^{1,2}$ is a differential pursuit of the (minus) second derivatives of functions $k_{1,2}^*$ at peak positions. The velocity of peaks drift is proportional to the first derivatives of these functions with the coefficients $1/b^{1,2}$. Therefore, velocity is proportional to the peak variance or more precise, to $\frac{\mathrm{var}}{t}$. Already such simple systems as (41) demonstrate various regimes of coevolution [46]

6 Simple Example of Arbitrary Complex Dynamics of Drift

Let X be a closed domain in R^n with nonempty interior. For a smooth vector field $v(x)$ in X we would like to construct such a Darwin's system (1) that drift (in logarithmic time $\tau = \ln t$) approximates dynamics defined by differential equation $\dot{x} = v(x)$. In order to consider this dynamics in X, some additional assumptions are needed. To guarantee positive invariance of X we can assume that there exists such $\varepsilon > 0$ that if $x \in X$ and $0 < \delta \le \varepsilon$ then $x + \delta v(x) \in X$. To consider function $v(x)$ in a vicinity of X we will use an arbitrary smooth continuation of $v(x)$ on R^n.

For any measure μ on X we use notations:

$$M_0(\mu) = \int_X \mu \,\mathrm{d}x, \quad M_1(\mu) = \int_X x\mu \,\mathrm{d}x, \quad M_2(\mu) = \int_X x^2\mu \,\mathrm{d}x\,.$$

Let us select the reproduction coefficient in the following form:

$$K(\mu)(x) = -(x - v(M_1(\mu)))^2 M_0(\mu) + C(\mu)\,, \tag{42}$$

where functional $C(\mu)$ is selected in such a way that $M_0(\mu)$ satisfies exactly equation $\dot{M}_0 = (1 - M_0)M_0$ if $\dot{\mu} = K(\mu)\mu$. Darwin's equation with the reproduction coefficient (42) gives for time derivative of M_0

$$\dot{M}_0 = \int_X \dot{\mu}(x)\,\mathrm{d}x = -M_2M_0 + 2(x, v(M_1))M_0 - v^2(M_1)M_0 + CM_0\,.$$

It is straightforward to check that for functional

$$C(\mu) = 1 - M_0 + M_2 - 2(M_1, v(M_1)) + M_0v^2(M_1)$$

dynamics of $M_0(\mu)$ satisfies the simple equation $\dot{M}_0 = (1-M_0)M_0$. Therefore, for positive initial condition after sufficiently long time the value of M_0 is arbitrarily closed to one. After some rearranging of coefficients, we get for time averages of $K(\mu)$:

$$\begin{aligned} \langle K(\mu)(x)\rangle_t &= \frac{1}{t}\int_0^t K(\mu(\tau))(x)\,\mathrm{d}\tau \\ &= -\left(x - \frac{\langle M_0v(M_1)\rangle_t}{\langle M_0\rangle_t}\right)^2 \langle M_0\rangle_t - \langle M_0v^2(M_1)\rangle_t + \frac{\langle M_0v(M_1)\rangle_t^2}{\langle M_0\rangle_t} + \langle C\rangle_t\,. \end{aligned} \tag{43}$$

This average reproduction coefficient achieves its maximum at point

$$x^* = \frac{\langle M_0v(M_1)\rangle_t}{\langle M_0\rangle_t}\,.$$

After sufficiently long time $x^* \approx \langle v(M_1)\rangle_t$, hence, for analysis of drift dynamics we have to study motion of the point $\langle v(M_1)\rangle_t$. By definition of time average, the velocity of this point in logarithmic time is

$$\frac{\mathrm{d}\langle v(M_1)\rangle_t}{\mathrm{d}\ln t} = v(M_1) - \frac{\langle v(M_1)\rangle_t}{t}\,.$$

For large t the second term tends to zero and we found that the time derivative of the reproduction coefficient maximizer x^* is $v(M_1)$: $\dot{x}^* = v(M_1)$ with arbitrarily chosen accuracy.

This is not yet an equation for peak motion. We need additional asymptotic identity $x^* \approx M_1$. It is not always true because it is possible that $\mathrm{supp}\mu \neq X$. Nevertheless, if at the initial moment $\mathrm{supp}\mu = X$ then for sufficiently large t $x^* \approx M_1$ because $\mu(t) = \mu(0)\exp(t\langle K(\mu)(x)\rangle_t)$ and almost all measure $\mu(t)$ is concentrated in an arbitrarily small vicinity of x^*. Therefore, $x^* \approx M_1$.

Finally, for drift dynamics we obtain equation: with an arbitrarily chosen accuracy in logarithmic time

$$\dot{x}^* = v(x^*)\,.$$

This simple example demonstrates that the drift of density peaks for Darwin's equations may be arbitrarily complex.

7 Conclusion

Darwin's equation demonstrate a mechanism of self-simplification of complex system. This mechanism, under the name "natural selection" was extracted from analysis of biological evolution by Charles Robert Darwin and Alfred Russel Wallace and published in 1859. Selection mechanism is based on specific separation of time: the support of distributions changes very slowly (small mutations) or cannot increase at al (inheritance).

Such a separation of time scales implies typical asymptotic behavior: supports of the ω-limit distributions are discrete. The asymptotic layer near the ω-limit distributions is drift of finite number of narrow density peaks. This drift becomes slower in time, but its dynamics in logarithmic time could be arbitrarily complex.

The equations for peak dynamics, the drift equations, (20,21,24) describe dynamics of the shapes of the peaks and their positions. For systems with small variability ("mutations") the drift equations (24) has the scaling property: the change of the intensity of mutations is equivalent to the change of the time scale.

Some further exact results of the mathematical selection theory can be found in [23,35,36]. Karev [33] recently developed an entropic description of limit behaviour of replicator systems.

There exists an important class of generalization of all selection theorems for distributions with vector space of values. In biological language this means that non-inherited properties are taken into account: distribution in size, age, space of birth and so on. The results are, essentially, the same: the Perron–Frobenius theorem and it generalizations allow to reduce the vector Darwin's systems back to the scalar optimality principle. The key role in this reduction plays the Birkhoff contraction theorem [2,16]

Many examples of Darwin's systems outside the theory of biological evolution in physics and other applications, such as weak turbulence or wave turbulence theory [53,54,38] or ecological applications [47], are already known. Nevertheless, this mechanism is, perhaps, still underestimated and we will meet them in many other areas.

References

1. G. D. Birkhoff. Dynamical systems. AMS Colloquium Publications, Providence, 1927.
2. G. Birkhoff, Extensions of Jentzsch's theorem, Trans. Amer. Math. Soc., 85, 219-227 (1957).
3. D. T. Bishop and C. Cannings. *A generalized war of attrition.* J. Theoret. Biol., 1978, Vol. 70, No. 1, 85-124.
4. I. M. Bomze. *Dynamical aspects of evolutionary stability.* Monatshefte für Mathematik, 1990, Vol. 110, 189-206.
5. I. M. Bomze. *Detecting all evolutionarily stable strategies.* J. Optim. Theory Appl., 1992, Vol. 75, No. 2, 313-329.

6. I. M. Bomze. *Uniform barriers and evolutionarily stable sets.* In: Game theory, experience, rationality. Foundations of social sciences, economics and ethics. In honor of John C. Harsanyi. Edited by Werner Leinfellner and Eckehart Khler. Vienna Circle Institute Yearbook, 5. Kluwer Academic Publishers, Dordrecht, 1998, 225-243.
7. I. M. Bomze. *Regularity vs. degeneracy in dynamics, games, and optimization: a unified approach to different aspects.* SIAM Review, 2002, Vol. 44, 394-414.
8. I. M. Bomze and R. Bürger. *Stability by mutation in evolutionary games, Evolutionary game theory in biology and economics* Games Econom. Behav., 1995, Vol. 11, No. 2, 146-172.
9. I. M. Bomze and B. M. Pötscher. Game Theoretical Foundations of Evolutionary Stability. Lecture Notes in Economics and Math. Systems 324, Springer, Berlin, 1989.
10. N. Bourbaki. Elements of mathematics – Integration I. Springer, Berlin – Heidelberg – New York, 2003.
11. Constantin, P., Foias, C., Nicolaenko, B., Temam, R., Integral manifolds and inertial manifolds for dissipative partial differential equations, Applied Math. Sci., 1988, Vol. 70 (Springer Verlag, New York).
12. R. Cressman. The Stability Concept of Evolutionary Game Theory. Lecture Notes in Biomathematics 94. Springer, Berlin, 1992.
13. Darwin, Ch., 1964, On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life: A Facsimile of the First Edition, Harvard.
14. O. Diekmann, M. Gyllenberg, and J. A. J. Metz. *Steady state analysis of structured population models.* Theor. Popul. Biol., 2003, Vol. 63, 309-338.
15. I. Eshel and E. Sansone. *Evolutionary and dynamic stability in continuous population games.* J. Math. Biology, 2003, Vol. 46, No. 5, 445-459.
16. S.P. Eveson, R.D. Nussbaum, Applications of the BirkhoffHopf theorem to the spectral theory of positive linear operators, Math. Proc. Cambridge Philos. Soc. 117 (1995) 491-512.
17. W. J. Ewens. Mathematical Population Genetics. Springer, Berlin, 1979.
18. M. Gardner. *On cellular automata, self-reproduction, the garden of Eden and the game of 'life'.* Scientific American, 1971, Vol. 224, No. 2, 112-115.
19. G. F. Gause. The struggle for existence. Williams and Wilkins, Baltimore, 1934. Online: http://www.ggause.com/Contgau.htm.
20. A. N. Gorban. *Dynamical systems with inheritance.* In: Some problems of community dynamics, R. G. Khlebopros (ed.). Institute of Physics RAS, Siberian Branch, Krasnoyarsk, 1980 [in Russian].
21. A. N. Gorban. Equilibrium encircling. Equations of chemical kinetics and their thermodynamic analysis. Nauka, Novosibirsk, 1984 [in Russian].
22. A. N. Gorban. *Order–disorder separation: Geometric revision,* Physica A, 2007, Vol. 374, 85-102.
23. A. N. Gorban, Selection Theorem for Systems with Inheritance, Math. Model. Nat. Phenom., 2 (4) (2007), 1-45. E-print: http://arxiv.org/abs/cond-mat/0405451
24. A.N. Gorban, I.V. Karlin, Invariant Manifolds for Physical and Chemical Kinetics, Lect. Notes Phys. 660, Springer, Berlin, Heidelberg, 2005.
25. A. N. Gorban and R. G. Khlebopros. Demon of Darwin: Idea of optimality and natural selection. Nauka (FizMatGiz), Moscow, 1988 [in Russian].

26. M. Gromov. *A dynamical model for synchronisation and for inheritance in micro-evolution: a survey of papers of A. Gorban.* The talk given in the IHES seminar, "Initiation to functional genomics: biological, mathematical and algorithmical aspects", Institut Henri Poincaré, November 16, 2000.
27. J. B. S. Haldane. The Causes of Evolution. Princeton Science Library, Princeton University Press, 1990.
28. P. Hammerstein. *Darwinian adaptation, population genetics and the streetcar theory of evolution.* J. Math. Biol., 1996, Vol. 34, 511-532.
29. G. Hardin. *The Competitive Exclusion Principle.* Science, 1960, 29 April, 1292-1297.
30. B. Hasselblatt and A. Katok (Eds.). Handbook of Dynamical Systems. Vol. 1A. Elsevier, Amsterdam, 2002.
31. J. Hofbauer, P. Schuster, and K. Sigmund. *A note on evolutionary stable strategies and game dynamics.* J. Theor. Biol., 1979, Vol. 81, 609-612.
32. J. Hofbauer and K. Sigmund. *Evolutionary game dynamics.* Bull. (New Series) American Math. Soc., 2003, Vol. 40, No. 4, 479-519.
33. Karev G.P., On mathematical theory of selection: continuous time population dynamics, Journal Of Mathematical Biology 60 (1) (2010), 107-129.
34. A. Katok and B. Hasselblat. Introduction to the Modern Theory of Dynamical Systems. Encyclopedia of Math. and its Applications, Vol. 54. Cambridge University Press, Cambridge, 1995.
35. O. A. Kuzenkov. *Weak solutions of the Cauchy problem in the set of Radon probability measures.* Differential Equations, 2000, Vol. 36, No. 11, 1676-1684.
36. O. A. Kuzenkov. *A dynamical system on the set of Radon probability measures.* Differential Equations, 1995, Vol. 31, No. 4, 549-554.
37. Levin, S.M., Community equilibria and stability, and an extension of the competitive exclusion principle, Am. Nat., 104 (939) (1970), pp. 413-423.
38. V. S. L'vov. Wave turbulence under parametric excitation applications to magnets. Springer, Berlin – Heidelberg, 1994.
39. R. MacArthur and R. Levins. *Competition, habitat selection, and character displacement in a patchy environment.* Proc. Nat. Acad. Sci., 1964, Vol. 51, pp. 1207-1210.
40. R. MacArthur and R. Levins. *The limiting similarity, convergence, and divergence of coexisting species.* Am. Nat., 1967, Vol. 101, No. 921, 377-385.
41. E. Mayr. Animal Species and Evolution. Harvard University Press, Cambridge, MA, 1963.
42. G. Meszena, M. Gyllenberg, L. Pasztor, and J. A. J. Metz. *Competitive exclusion and limiting similarity: A unified theory.* Theoretical Population Biology, 2006, Vol. 69, 68-87.
43. G. Meszena and J. A. J. Metz. *Species diversity and population regulation: the importance of environmental feedback dimensionality.* International Institute for Applied SystemsAnalysis, Interim Report IR-99-045, 1999.
44. J. Oechssler and F. Riedel. *Evolutionary dynamics on infinite strategy spaces.* Economic Theory, 2001, Vol. 17, No. 1, 141-162.
45. J. Oechssler and F. Riedel. *On the Dynamic Foundation of Evolutionary Stability in Continuous Models.* Journal of Economic Theory, 2002, Vol. 107, 223-252.
46. L. I. Rozonoer and E. I. Sedyh. *On the mechanisms of of evolution of self-reproduction systems. 1, 2, 3.* Automation and Remote Control, 1979, Vol. 40, 243-251; 419-429; 741-749.

47. F. N. Semevsky and S. M. Semenov. Mathematical modeling of ecological processes. Gidrometeoizdat, Leningrad, 1982 [in Russian].
48. J. M. Smith and G. R. Price. *The Logic of Animal Conflict.* Nature, 1973, Vol. 246, No. 5427, 15-18.
49. J. M. Smith. Evolution and the theory of games. Cambridge Univ. Press, Cambridge, 1982.
50. P. Taylor. *Evolutionary stable strategies and game dynamics.* J. Appl. Probability, 1979, Vol. 16, 76-83.
51. G. T. Vickers and C. Cannings. *Patterns of ESSs. I, II.* J. Theoret. Biol., 1988, Vol. 132, No. 4, 387-408, 409-420.
52. V. Volterra. Lecons sur la théorie mathematique de la lutte pour la vie. Gauthier-Villars, Paris, 1931.
53. V. E. Zakharov, V. S. L'vov, and S. S. Starobinets. *Turbulence of spin-waves beyond threshold of their parametric-excitation.* Uspekhi Fizicheskikh Nauk, 1974, Vol. 114, No. 4, 609-654; English translation: Sov. Phys. – Usp., 1975, Vol. 17, No. 6, 896-919.
54. V. E. Zakharov, V. S. L'vov, and G. E. Falkovich. Kolmogorov spectra of turbulence. Vol. 1. Wave Turbulence. Springer, Berlin, 1992.